**AI Glossary**

**Accuracy**

Accuracy is a scoring system in binary classification (i.e., determining if an answer or output is correct or not) and is calculated as (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives).

**Actionable Intelligence**

Information you can leverage to support decision making.

**Anaphora**

In linguistics, an anaphora is a reference to a noun by way of a pronoun. For example, in the sentence, “While John didn’t like the appetizers, he enjoyed the entrée,” the word “**he” is an anaphora.**

**Annotation**

The process of tagging language data by identifying and flagging grammatical, semantic or phonetic elements in language data.

**Artificial General Intelligence (or "AGI")**

Artificial General Intelligence (AGI) refers to an AI system that possesses a wide range of cognitive abilities, much like humans, enabling them to learn, reason, adapt to new situations, and devise creative solutions across various tasks and domains, rather than being limited to specific tasks as narrow AI systems are.

**Artificial Neural Network (ANN)**

Commonly referred to as a neural network, this system consists of a collection of nodes/units that loosely mimics the processing abilities of the human brain.

**Artificial Intelligence (or “AI”)**

The simulation of human intelligence in machines that are programmed to think and learn like humans. Example: A self-driving car that can navigate and make decisions on its own using AI technology.

**AI ethics**

Refers to the issues that AI stakeholders such as engineers and government officials must consider to ensure that the technology is developed and used responsibly. This means adopting and implementing systems that support a safe, secure, unbiased, and environmentally friendly approach to artificial intelligence.

**Algorithm**

An algorithm is a sequence of rules given to an AI machine to perform a task or solve a problem. Common algorithms include classification, regression, and clustering.

**Alignment**

Tweaking an AI to better produce the desired outcome. This can refer to anything from moderating content to maintaining positive interactions toward humans.

**Auto-classification**

The application of machine learning, natural language processing (NLP), and other AI-guided techniques to automatically classify text in a faster, more cost-effective, and more accurate manner.

**Auto-complete**

Auto-complete is a search functionality used to suggest possible queries based on the text being used to compile a search query.

**Autonomous agents**

An AI model that have the capabilities, programming and other tools to accomplish a specific task. A self-driving car is an autonomous agent, for example, because it has sensory inputs, GPS and driving algorithms to navigate the road on its own. [Stanford researchers](https://arxiv.org/pdf/2304.03442) have shown that autonomous agents can develop their own cultures, traditions and shared language.

**Benchmarking**

Benchmarking is the process of evaluating and comparing products or systems using standardized tests to gauge performance and capabilities.

**BERT (aka Bidirectional Encoder Representation from Transformers)**

Google’s technology. A large scale pretrained model that is first trained on very large amounts of unannotated data. The model is then transferred to an NLP task where it is fed another smaller task-specific dataset which is used to fine-tune the final model.

**Bias**

In regards to large language models, errors resulting from the training data. This can result in falsely attributing certain characteristics to certain races or groups based on stereotypes.

**Cataphora**

In linguistics, a cataphora is a reference placed before any instance of the noun it refers to. For example, in the sentence, “Though he enjoyed the entrée, John didn’t like the appetizers,” the word “he” is a cataphora.

**Categorization**

Categorization is a natural language processing function that assigns a category to a document.

**Category**

A category is a label assigned to a document in order to describe the content within said document.

**Category Trees**

Enables you to view all of the rule-based categories in a collection. Used to create categories, delete categories, and edit the rules that associate documents with categories. Is also called a taxonomy, and is arranged in a hierarchy.

**Chatbot**

A user-friendly interface that allows the user to ask questions and receive answers. Depending on the backend system that fuels the chatbot, it can be as basic as pre-written responses to a fully conversational AI that automates issue resolution.

**Classification**

Techniques that assign a set of predefined categories to open-ended text to be used to organize, structure, and categorize any kind of text – from documents, medical records, emails, files, within any application and across the web or social media networks.

**Collective Learning**

Collective learning is an AI training approach that leverages diverse skills and knowledge across multiple models to achieve more powerful and robust intelligence.

**Co-occurrence**

A co-occurrence commonly refers to the presence of different elements in the same document. It is often used in business intelligence to heuristically recognize patterns and guess associations between concepts that are not naturally connected (e.g., the name of an investor often mentioned in articles about startups successfully closing funding rounds could be interpreted as the investor is particularly good at picking his or her investments.).

**Cognitive computing**

mimicking human thought processes such as pattern recognition and learning. Marketing teams sometimes use this term to eliminate the sci-fi mystique of AI.

**Cognitive Map**

A mental representation (otherwise known as a mental palace) which serves an individual to acquire, code, store, recall, and decode information about the relative locations and attributes of phenomena in their environment.

**Completions**

The output from a generative prompt.

**Composite AI**

The combined application of different AI techniques to improve the efficiency of learning in order to broaden the level of knowledge representations and, ultimately, to solve a wider range of business problems in a more efficient manner.

**Computational Linguistics (Text Analytics, Text Mining)**

Computational linguistics is an interdisciplinary field concerned with the computational modeling of natural language.

**Computational Semantics (Semantic Technology)**

Computational semantics is the study of how to automate the construction and reasoning of meaning representations of natural language expressions.

**Content**

Individual containers of information — that is, documents — that can be combined to form training data or generated by Generative AI.

**Content Enrichment or Enrichment**

The process of applying advanced techniques such as machine learning, AI, and language processing to automatically extract meaningful information from your text-based documents.

**Controllability**

Controllability is the ability to understand, regulate, and manage an AI system's decision-making process, ensuring its accuracy, safety, and ethical behavior, and minimizing the potential for undesired consequences.

**Controlled Vocabulary**

A controlled vocabulary is a curated collection of words and phrases that are relevant to an application or a specific industry. These elements can come with additional properties that indicate both how they behave in common language and what meaning they carry, in terms of topic and more.

While the value of a controlled vocabulary is similar to that of taxonomy, they differ in that the nodes in taxonomy are only labels representing a category, while the nodes in a controlled vocabulary represent the words and phrases that must be found in a text.

**Conversational AI**

Used by developers to build conversational user interfaces, chatbots and virtual assistants for a variety of use cases. They offer integration into chat interfaces such as messaging platforms, social media, SMS and websites. A conversational AI platform has a developer API so third parties can extend the platform with their own customizations.

**Convolutional Neural Networks (CNN)**

A deep learning class of neural networks with one or more layers used for image recognition and processing.

**Corpus**

The entire set of language data to be analyzed. More specifically, a corpus is a balanced collection of documents that should be representative of the documents an NLP solution will face in production, both in terms of content as well as distribution of topics and concepts.

**Custom/Domain Language model**

A model built specifically for an organization or an industry – for example Insurance.

**Data Augmentation**

Data Augmentation is a technique used to artificially increase the size and diversity of a training set by creating modified copies of the existing data. It involves making minor changes such as flipping, resizing, or adjusting the brightness of images, to enhance the dataset and prevent models from overfitting.

**Data Discovery**

The process of uncovering data insights and getting those insights to the users who need them, when they need them.

**Data Drift**

Data Drift occurs when the distribution of the input data changes over time; this is also known as covariate shift.

**Data Extraction**

Data extraction is the process of collecting or retrieving disparate types of data from a variety of sources, many of which may be poorly organized or completely unstructured.

**Data Ingestion**

The process of obtaining disparate data from multiple sources, restucturing it, and importing it into a common format or repository to make it easy to utilize.

**Data Labelling**

A technique through which data is marked to make objects recognizable by machines. Information is added to various data types (text, audio, image and video) to create metadata used to train AI models.

**Data Scarcity**

The lack of data that could possibly satisfy the need of the system to increase the accuracy of predictive analytics.

**Deep Learning**

Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. In other words, deep learning models can learn to classify concepts from images, text or sound.

**Deterministic Model**

A deterministic model follows a specific set of rules and conditions to reach a definite outcome, operating on a cause-and-effect basis.

**Did You Mean (DYM)**

“Did You Mean” is an NLP function used in search applications to identify typos in a query or suggest similar queries that could produce results in the search database being used.

**Diffusion**

A method of machine learning that takes an existing piece of data, like a photo, and adds random noise. Diffusion models train their networks to re-engineer or recover that photo.

**Disambiguation**

Disambiguation, or word-sense disambiguation, is the process of removing confusion around terms that express more than one meaning and can lead to different interpretations of the same string of text.

**Discriminative Model**

Discriminative models are algorithms designed to directly model and learn the boundary between different classes or categories in a dataset.

**Domain Knowledge**

The experience and expertise your organization has acquired over time.

**Edge model**

A model that includes data typically outside centralized cloud data centers and closer to local devices or individuals — for example, wearables and Internet of Things (IoT) sensors or actuators.

**Embedding**

A set of data structures in a large language model (LLM) of a body of content where a high-dimensional vector represents words. This is done so data is more efficiently processed regarding meaning, translation and generation of new content.

**Emergent behavior**

Emergent behavior, also called emergence, is when an AI system shows unpredictable or unintended capabilities.

**Emotion AI (aka Affective Computing)**

AI to analyze the emotional state of a user (via computer vision, audio/voice input, sensors and/or software logic). It can initiate responses by performing specific, personalized actions to fit the mood of the customer.

**End-to-end learning, or E2E**

A deep learning process in which a model is instructed to perform a task from start to finish. It's not trained to accomplish a task sequentially but instead learns from the inputs and solves it all at once.

**Entity**

An entity is any noun, word or phrase in a document that refers to a concept, person, object, abstract or otherwise (e.g., car, Microsoft, New York City). Measurable elements are also included in this group (e.g., 200 pounds, 14 fl. oz.)

**Enterprise AI**

Enterprise AI refers to the strategic integration and deployment of AI within an organizational framework to enhance various business processes, decision-making, and overall operational efficiency.

**Environmental, Social, and Governance (ESG)**

An acronym initially used in business and government pertaining to enterprises’ societal impact and accountability; reporting in this area is governed by a set of binding and voluntary regulatory reporting.

**ETL (Entity Recognition, Extraction)**

Entity extraction is an NLP function that serves to identify relevant entities in a document.

**Ethical considerations**

An awareness of the ethical implications of AI and issues related to privacy, data usage, fairness, misuse and other safety issues.

**Explainable AI/Explainability**

An AI approach where the performance of its algorithms can be trusted and easily understood by humans. Unlike black-box AI, the approach arrives at a decision and the logic can be seen behind its reasoning and results.

**Extensibility**

Extensibility in AI refers to the ability of AI systems to expand their capabilities to new domains, tasks, and datasets without needing full retraining or major architectural changes.

**Extraction or Keyphrase Extraction**

Multiple words that describe the main ideas and essence of text in documents.

**Extractive Summarization**

Identifies the important information in a text and groups the text fragments together to form a concise summary. Also see **Generative Summarization**

**F-score (F-measure, F1 measure)**

An F-score is the harmonic mean of a system’s precision and recall values. It can be calculated by the following formula: 2 x [(Precision x Recall) / (Precision + Recall)]. Criticism around the use of F-score values to determine the quality of a predictive system is based on the fact that a moderately high F-score can be the result of an imbalance between precision and recall and, therefore, not tell the whole story. On the other hand, systems at a high level of accuracy struggle to improve precision or recall without negatively impacting the other.

Critical (risk) applications that value information retrieval more than accuracy (i.e., producing a large number of false positives but virtually guaranteeing that all the true positives are found) can adopt a different scoring system called F2 measure, where recall is weighed more heavily. The opposite (precision is weighed more heavily) is achieved by using the F0.5 measure.

**Few-shot learning**

In contrast to traditional models, which require many training examples, few-shot learning uses only a small number of training examples to generalize and produce worthwhile output.

**Fine-tuned model**

A model focused on a specific context or category of information, such as a topic, industry or problem set

**Fine-tuning**

Improving an existing, pretrained model through additional training with new, context- or task-specific data.

**Foom**

Also known as fast takeoff or hard takeoff. The concept that if someone builds an AGI that it might already be too late to save humanity.

**Foundational model**

A baseline model used for a solution set, typically pretrained on large amounts of data using self-supervised learning. Applications or other models are used on top of foundational models — or in fine-tuned contextualized versions. Examples include BERT, GPT-n, Llama, DALL-E, etc.

**Generalized model**

A model that does not specifically focus on use cases or information.

**Generation**

Generation is the ability of a generative model to create brand new, original content such as text, images, audio or video from scratch.

**Generative AI (GenAI)**

AI techniques that learn from representations of data and model artifacts to generate new artifacts.

**Generative adversarial networks (or "GANs")**

GANs are a powerful type of neural network capable of generating new, never-seen-before data that closely resembles the training data.

**Generative Pre-Trained Transformer**

Generative pre-trained transformers (GPT) are neural network models trained on large datasets in an unsupervised manner to generate text.

**Generative Summarization**

Using LLM functionality to take text prompt inputs like long form chats, emails, reports, contracts, policies, etc and distilling them down to their core content, generating summaries from the text prompts for quick comprehension. Thus using pre-trained language models and context understanding to produce concise, accurate and relevant summaries.

**Grounding**

The ability of generative applications to map the factual information contained in a generative output or completion. It links generative applications to available factual sources — for example, documents or knowledge bases — as a direct citation, or it searches for new links.

**Guardrails**

Guardrails refers to restrictions and rules placed on AI systems to make sure that they handle data appropriately and don't generate unethical content.

**Hallucinations**

Made up data presented as fact in generated text that is plausible but are, in fact, inaccurate or incorrect. Made up data that include fabricated, inaccurate or misaligned references or sources that are presented as fact in generated text.

**Hybrid AI**

Hybrid AI is any artificial intelligence technology that combines multiple AI methodologies. In NLP, this often means that a workflow will leverage both symbolic and machine learning techniques.

**Hyperparameters**

These are adjustable model parameters that are tuned in order to obtain optimal performance of the model.

**Inference Engine**

A component of a [expert] system that applies logical rules to the knowledge base to deduce new or additional information.

**Insight Engines**

An insight engine, also called cognitive search or enterprise knowledge discovery. It applies relevancy methods to describe, discover, organize and analyze data. It combines search with AI capabilities to provide information for users and data for machines. The goal of an insight engine is to provide timely data that delivers actionable intelligence.

**Instruction-tuning**

Instruction-tuning is an approach where a pre-trained model is adapted to perform specific tasks by providing a set of guidelines or directives that outline the desired operation.

**Intelligent Document Processing (IDP) or Intelligent Document Extraction and Processing (IDEP)**

This is the ability to automatically read and convert unstructured and semi-structured data, identify usable data and extract it, then leveraged it via automated processes. IDP is often an enabling technology for Robotic Process Automation (RPA) tasks.

**Interpretability**

Interpretability refers to how inherently understandable or explainable an AI model is based on its architecture, logic, and behavior.

**Knowledge Engineering**

A method for helping computers replicate human-like knowledge. Knowledge engineers build logic into knowledge-based systems by acquiring, modeling and integrating general or domain-specific knowledge into a model.

**Knowledge Graph**

A knowledge graph is a graph of concepts whose value resides in its ability to meaningfully represent a portion of reality, specialized or otherwise. Every concept is linked to at least one other concept, and the quality of this connection can belong to different classes

The interpretation of every concept is represented by its links. Consequently, every node is the concept it represents only based on its position in the graph (e.g., the concept of an apple, the fruit, is a node whose parents are “apple tree”, “fruit”, etc.). Advanced knowledge graphs can have many properties attached to a node including the words used in language to represent a concept (e.g., “apple” for the concept of an apple), if it carries a particular sentiment in a culture (“bad”, “beautiful”) and how it behaves in a sentence.

**Knowledge graphs**

Machine-readable data structures representing knowledge of the physical and digital worlds and their relationships. Knowledge graphs adhere to the graph model — a network of nodes and links.

**Knowledge Model**

A process of creating a computer interpretable model of knowledge or standards about a language, domain, or process(es). It is expressed in a data structure that enables the knowledge to be stored in a database and be interpreted by software.

**Knowledged Based AI**

Knowledge-based systems (KBs) are a form of artificial intelligence (AI) designed to capture the knowledge of human experts to support decision-making and problem-solving.

**Knowledge Generation**

Knowledge generation involves training models on extensive datasets, allowing them to analyze data, discover patterns, and craft new insights.

**K-Shot Learning**

K-shot learning is a machine learning approach where models learn from only k labeled examples per class, where k is a small number like 1-5.

**Labelled Data** see Data Labelling.

**LangOps (Language Operations)**

The workflows and practices that support the training, creation, testing, production deployment and ongoing curation of language models and natural language solutions.

**Language Data**

Language data is data made up of words; it is a form of unstrcutured data. This is qualitative data and also known as text data, but simply it refers to the written and spoken words in language.

**Large Language Models (LLM)**

A supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

**Latency**

Latency refers to the time delay between when an AI system receives an input and generates the corresponding output.

**Lemma**

The base form of a word representing all its inflected forms.

**Lexicon**

Knowledge of all of the possible meanings of words, in their proper context; is fundamental for processing text content with high precision.

**Limited memory**

Limited memory is a type of AI system that receives knowledge from real-time events and stores it in the database to make better predictions.

**Linked Data**

Linked data is an expression that informs whether a recognizable store of knowledge is connected to another one. This is typically used as a standard reference. For instance, a knowledge graph in which every concept/node is linked to its respective page on Wikipedia.

**Low-code**

Low-code is a visual approach to software development that enables faster delivery of applications through minimal hand-coding.

**Machine Learning (ML)**

Machine learning is the study of computer algorithms that can improve automatically through experience and the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as “training data,” in order to make predictions or decisions without being explicitly programmed to do so. In NLP, ML-based solutions can quickly cover the entire scope of a problem (or, at least of a corpus used as sample data), but are demanding in terms of the work required to achieve production-grade accuracy.

**Metacontext and metaprompt**

Foundational instructions on how to train the way in which the model should behave.

**Metadata**

Data that describes or provides information about other data.

**Model**

A machine learning model is the artifact produced after an ML algorithm has processed the sample data it was fed during the training phase. The model is then used by the algorithm in production to analyze text (in the case of NLP) and return information and/or predictions.

**Model Chaining**

Model chaining is a technique in data science where multiple machine learning models are linked in a sequence to make predictions or analyzations.

**Model Drift**

Model drift is the decay of models’ predictive power as a result of the changes in real world environments. It is caused due to a variety of reasons including changes in the digital environment and ensuing changes in relationship between variables. An example is a model that detects spam based on email content and then the content used in spam was changed.

**Model Parameter**

These are parameters in the model that are determined by using the training data. They are the fitted/configured variables internal to the model whose value can be estimated from data. They are required by the model when making predictions. Their values define the capability and fit of the model.

**Morphological Analysis**

Breaking a problem with many known solutions down into its most basic elements or forms, in order to more completely understand them. Morphological analysis is used in general problem solving, linguistics and biology.

**Multi-hop Reasoning**

Multi-hop is a term often used in natural language processing and, more specifically, machine reading comprehension tasks. It refers to the process by which an AI model retrieves answers to questions by connecting multiple pieces of information present in a given text or across various sources and systems, rather than directly extracting the information from a single passage.

**Multimodal models and modalities**

Language models that are trained on and can understand multiple data types, such as words, images, audio and other formats, resulting in increased effectiveness in a wider range of tasks

**Multitask prompt tuning (MPT)**

An approach that configures a prompt representing a variable — that can be changed — to allow repetitive prompts where only the variable changes.

**N-Shot Learning**

Zero/Single/Few shot learning are variations of the same concept – providing a model with little or no training data to classify new data and guide predictions. A “shot” represents a single training example. Fun fact: Within the GPT prompt, you can ask for “N” examples to improve the accuracy of the response.

**Natural Language Processing**

A subfield of artificial intelligence and linguistics, natural language processing is focused on the interactions between computers and human language. More specifically, it focuses on the ability of computers to read and analyze large volumes of unstructured language data (e.g., text).

**Natural Language Understanding**

A subset of natural language processing, natural language understanding is focused on the actual computer comprehension of processed and analyzed unstructured language data. This is enabled via semantics.

**Natural Language Generation - NLG**

Solutions that automatically convert structured data, such as that found in a database, an application or a live feed, into a text-based narrative. This makes the data easier for users to access by reading or listening, and therefore to comprehend.

**Natural Language Query- NLQ**

A natural language input that only includes terms and phrases as they occur in spoken language (i.e. without non-language characters).

**Natural Language Technology - NLT**

A subfield of linguistics, computer science and artificial intelligence (AI) dealing with Natural Language Processing (NLP), Natural Language Undestanding (NLU), and Natural Language Generation (NLG).

**Neural Network**

A machine learning model inspired by the human brain's structure and function that's composed of layers of interconnected nodes or "neurons." Example: A neural network that can recognize handwritten digits with high accuracy.

**OpenAI**

The organization that developed ChatGPT. More broadly speaking, OpenAI is a research company that aims to develop and promote friendly AI responsibly. Example: OpenAI's GPT-3 model is one of the largest and most powerful language models available for natural language processing tasks.

**Ontology**

An ontology is similar to a taxonomy, but it enhances its simple tree-like classification structure by adding properties to each node/element and connections between nodes that can extend to other branches. These properties are not standard, nor are they limited to a predefined set. Therefore, they must be agreed upon by the classifier and the user.

**Optimization**

The process of adjusting the parameters of a model to minimize a loss function that measures the difference between the model's predictions and the true values. Example: Optimizing a neural network's parameters using a gradient descent algorithm to minimize the error between the model's predictions and the true values.

**Overfitting**

A problem that occurs when a model is too complex, performing well on the training data but poorly on unseen data. Example: A model that has memorized the training data instead of learning general patterns and thus performs poorly on new data.

**Paperclips**

The Paperclip Maximiser theory, coined by philosopher Nick Boström of the University of Oxford, is a hypothetical scenario where an AI system will create as many literal paperclips as possible. In its goal to produce the maximum amount of paperclips, an AI system would hypothetically consume or convert all materials to achieve its goal. This could include dismantling other machinery to produce more paperclips, machinery that could be beneficial to humans. The unintended consequence of this AI system is that it may destroy humanity in its goal to make paperclips.

**Parameters**

A set of numerical weights representing neural connections or other aspects in an AI model with values that are determined by training. Large language models (LLMs) can have billions of parameters.

**Parameter-efficient Fine-tuning (or "PEFT")**

Parameter-Efficient Fine-Tuning, also known as PEFT, is an approach that helps you improve the performance of large AI models while optimizing for resources like time, energy, and computational power. To do this, PEFT focuses on adjusting a small number of key parameters while preserving most of the pretrained model's structure.

**Parsing**

Identifying the single elements that constitute a text, then assigning them their logical and grammatical value.

**Part-of-Speech Tagging**

A Part-of-Speech (POS) tagger is an NLP function that identifies grammatical information about the elements of a sentence. Basic POS tagging can be limited to labeling every word by grammar type, while more complex implementations can group phrases and other elements in a clause, recognize different types of clauses, build a dependency tree of a sentence, and even assign a logical function to every word (e.g., subject, predicate, temporal adjunct, etc.).

**Post Edit Machine Translation** - **PEMT**

Solution allows a translator to edit a document that has already been machine translated. Typically, this is done sentence-by-sentence using a specialized computer-assisted-translation application.

**Plugins**

A software component or module that extends the functionality of an LLM system into a wide range of areas, including travel reservations, e-commerce, web browsing and mathematical calculations.

**Post-processing**

Procedures that can include various pruning routines, rule filtering, or even knowledge integration. All these procedures provide a kind of symbolic filter for noisy and imprecise knowledge derived by an algorithm.

**Pre-processing**

A step in the data mining and data analysis process that takes raw data and transforms it into a format that can be understood and analyzed by computers. Analyzing structured data, like whole numbers, dates, currency and percentages is straigntforward. However, unstructured data, in the form of text and images must first be cleaned and formatted before analysis.

**Precision**

Given a set of results from a processed document, precision is the percentage value that indicates how many of those results are correct based on the expectations of a certain application. It can apply to any class of a predictive AI system such as search, categorization and entity recognition.

For example, say you have an application that is supposed to find all the dog breeds in a document. If the application analyzes a document that mentions 10 dog breeds but only returns five values (all of which are correct), the system will have performed at 100% precision. Even if half of the instances of dog breeds were missed, the ones that were returned were correct.

**Predictive analytics**

Is a type of analytics that uses technology to predict what will happen in a specific time frame based on historical data and patterns.

**Prescriptive analytics**

Is a type of analytics that uses technology to analyze data for factors such as possible situations and scenarios, past and present performance, and other resources to help organizations make better strategic decisions.

**Pretrained model**

A model trained to accomplish a task — typically one that is relevant to multiple organizations or contexts. Also, a pretrained model can be used as a starting point to create a fine-tuned contextualized version of a model, thus applying transfer learning.

**Pretraining**

The first step in training a foundation model, usually done as an unsupervised learning phase. Once foundation models are pretrained, they have a general capability. However, foundation models need to be improved through fine-tuning to gain greater accuracy.

**Probabilistic Model**

A probabilistic AI model makes decisions based on probabilities or likelihoods.

**Prompt**

A phrase or individual keywords used as input for GenAI.

**Prompt chaining**

An approach that uses multiple prompts to refine a request made by a model.

**Prompt Engineering**

The craft of designing and optimizing user requests to an LLM or LLM-based chatbot to get the most effective result, often achieved through significant experimentation.

**Question & Answer (Q&A)**

An AI technique that allows users to ask questions using common everyday language and receive the correct response back. With the advent of large language models (LLMs), question and answering has evolved to let users ask questions using common everyday language and use Retrieval Augmented Generation (RAG) approaches to generate a complete answer from the text fragments identified in the target document or corpus.

**Quantum computing**

Is the process of using quantum-mechanical phenomena such as entanglement and superposition to perform calculations. Quantum machine learning uses these algorithms on quantum computers to expedite work because it performs much faster than a classic machine learning program and computer.

**Random Forest**

A supervised machine learning algorithm that grows and combines multiple decision trees to create a “forest.” Used for both classification and regression problems in R and Python.

**Reasoning**

AI reasoning is the process by which artificial intelligence systems solve problems, think critically, and create new knowledge by analyzing and processing available information, allowing them to make well-informed decisions across various tasks and domains.

**Recall**

Given a set of results from a processed document, recall is the percentage value that indicates how many correct results have been retrieved based on the expectations of the application. It can apply to any class of a predictive AI system such as search, categorization and entity recognition.

For example, say you have an application that is supposed to find all the dog breeds in a document. If the application analyzes a document that mentions 10 dog breeds but only returns five values (all of which are correct), the system will have performed at 50% recall.

**Recursive Prompting**

Recursive prompting is a strategy for guiding AI models like OpenAI's GPT-4 to produce higher-quality output. It involves providing the model with a series of prompts or questions that build upon previous responses, refining both the context and the AI's understanding to achieve the desired result.

**Recurrent Neural Networks (RNN)**

A neural network model commonly used in natural language process and speech recognition allowing previous outputs to be used as inputs.

**Reinforcement learning**

A machine learning (ML) training method that rewards desired behaviors or punishes undesired ones.

**Reinforcement learning with human feedback (RLHF)**

A ML algorithm that learns how to perform a task by receiving feedback from a human.

**Relations**

The identification of relationships is an advanced NLP function that presents information on how elements of a statement are related to each other. For example, “John is Mary’s father” will report that John and Mary are connected, and this datapoint will carry a link property that labels the connection as “family” or “parent-child.”

**Responsible AI**

Responsible AI is a broad term that encompasses the business and ethical choices associated with how organizations adopt and deploy AI capabilities. Generally, Responsible AI looks to ensure Transparent (Can you see how an AI model works?); Explainable (Can you explain why a specific decision in an AI model was made?); Fair (Can you ensure that a specific group is not disadvantaged based on an AI model decision?); and Sustainable (Can the development and curation of AI models be done on an environmentally sustainable basis?) use of AI.

**Retrieval Augmented Generation (RAG)**

Retrieval-augmented generation (RAG) is an AI technique for improving the quality of LLM-generated responses by including trusted sources of knowledge, outside of the original training set, to improve the accuracy of the LLM’s output. Implementing RAG in an LLM-based question answering system has benefits: 1) assurance that an LLM has access to the most current, reliable facts, 2) reduce hallucinations rates, and 3) provide source attribution to increase user trust in the output.

**ROAI**

Return on Artificial Intelligence is an abbreviation for return on investment on an AI-specific initiative or investment.

**Rules-based Machine Translation (RBMT)**

Considered the “Classical Approach” of machine translation it is based on linguistic information about source and target that allow words to have different meaning depending on the context.

**SAO (Subject-Action-Object)**

Subject-Action-Object (SAO) is an NLP function that identifies the logical function of portions of sentences in terms of the elements that are acting as the subject of an action, the action itself, the object receiving the action (if one exists), and any adjuncts if present. Learn more about:

**Self-supervised learning**

An approach to ML in which labeled data is created from the data itself. It does not rely on historical outcome data or external human supervisors that provide labels or feedback.

**Semantic Network**

A form of knowledge representation, used in several natural language processing applications, where concepts are connected to each other by semantic relationship.

**Semantic Search**

The use of natural language technologies to improve user search capabilities by processing the relationship and underlying intent between words by identifying concepts and entities such as people and organizations are revealed along with their attributes and relationships.

**Semantics**

Semantics is the study of the meaning of words and sentences. It concerns the relation of linguistic forms to non-linguistic concepts and mental representations to explain how sentences are understood by the speakers of a language.

**Semi-structured Data**

Data that is stuctured in some way but does not obey the tablular structure of traditional databases or other conventional data tables most commonly organized in rows and columns. Attributes of the data are different even though they may be grouped together. A simple example is a form; a more advanced example is a object database where the data is represented in the form of objects that are related (e.g. automobile make relates to model relates to trim level).

**Sentiment**

Sentiment is the general disposition expressed in a text.

**Sentiment Analysis**

Sentiment analysis is an NLP function that identifies the sentiment in text. This can be applied to anything from a business document to a social media post. Sentiment is typically measured on a linear scale (negative, neutral or positive), but advanced implementations can categorize text in terms of emotions, moods, and feelings.

**Sequence Modeling**

A subfield of NLP that focuses on modeling sequential data such as text, speech, or time series data. Example: A sequence model that can predict the next word in a sentence or generate coherent text.

**Similarity (and Correlation)**

Similarity is an NLP function that retrieves documents similar to a given document. It usually offers a score to indicate the closeness of each document to that used in a query. However, there are no standard ways to measure similarity. Thus, this measurement is often specific to an application versus generic or industry-wide use cases.

**Simple Knowledge Organization System (SKOS)**

A common data model for knowledge organization systems such as thesauri, classification schemes, subject heading systems, and taxonomies.

**Specialized corpora**

A focused collection of information or training data used to train an AI. Specialized corpora focuses on an industry — for example, banking, Insurance or health — or on a specific business or use case, such as legal documents.

**Speech Analytics**

The process of analyzing recordings or live calls with speech recognition software to find useful information and provide quality assurance. Speech analytics software identifies words and analyzes audio patterns to detect emotions and stress in a speaker’s voice.

**Speech Recognition**

Speech recognition or automatic speech recognition (ASR), computer speech recognition, or speech-to-text, enables a software program to process human speech into a written/text format.

**Stable Diffusion**

Stable diffusion is an artificial intelligence system that uses deep learning to generate images from text prompts.

**Stacking**

Stacking is a technique in AI that combines multiple algorithms to enhance overall performance. By blending the strengths of various AI models, stacking compensates for each model's weaknesses and achieves a more accurate and robust output in diverse applications, such as image recognition and natural language processing.

**Steerability**

AI steerability refers to the ability to guide or control an AI system's behavior and output according to human intentions or specific objectives. This involves designing AI models with mechanisms that understand and adhere to the preferences provided by users, while avoiding unintended or undesirable outcomes. Improving steerability requires ongoing research and refinement, including techniques like fine-tuning, rule-based systems, and implementing additional human feedback loops during AI development.

**Strong AI**

Strong AI refers to machines possessing generalized intelligence and capabilities on par with human cognition.

**Structured Data**

Structured data is the data which conforms to a specific data model, has a well-defined structure, follows a consistent order and can be easily accessed and used by a person or a computer program. Structured data are usually stored in rigid schemas such as databases.

**Stochastic Parrot**

Stochastic parrots are AI systems that use statistics to convincingly generate human-like text, while lacking true semantic understanding behind the word patterns.

**Style transfer**

The ability to adapt the style of one image to the content of another, allowing an AI to interpret the visual attributes of one image and use it on another. For example, taking the self-portrait of Rembrandt and re-creating it in the style of Picasso.

**Summarization (Text)**

Text summarization is the process of creating a short, accurate, and fluent summary of a longer text document. The goal is to reduce the size of the text while preserving its important information and overall meaning. There are two main types of text summarization:

Extractive Summarization

Generative Summarization also know as Abstractive Summarization

**Supervised learning**

An ML algorithm in which the computer is trained using labeled data or ML models trained through examples to guide learning.

**Symbolic AI**

Add to Symboilic Methodology parthethetically so it looks like this: “Symbolic Methodology (Symbolic AI)”

**Symbolic Methodology**

A symbolic methodology is an approach to developing AI systems for NLP based on a deterministic, conditional approach. In other words, a symbolic approach designs a system using very specific, narrow instructions that guarantee the recognition of a linguistic pattern. Rule-based solutions tend to have a high degree of precision, though they may require more work than ML-based solutions to cover the entire scope of a problem, depending on the application.

**Syntax**

The arrangement of words and phrases in a specific order to create meaning in language. If you change the position of one word, it is possible to change the context and meaning.

**Tagging**

See Parts-of-Speech Tagging (aka POS Tagging).

**Taxonomy**

A taxonomy is a predetermined group of classes of a subset of knowledge (e.g., animals, drugs, etc.). It includes dependencies between elements in a “part of” or “type of” relationship, giving itself a multi-level, tree-like structure made of branches (the final node or element of every branch is known as a leaf). This creates order and hierarchy among knowledge subsets.

Companies use taxonomies to more concisely organize their documents which, in turn, enables internal or external users to more easily search for and locate the documents they need. They can be specific to a single company or become de-facto languages shared by companies across specific industries.

**Temperature**

A parameter that controls the degree of randomness or unpredictability of the LLM output. A higher value means greater deviation from the input; a lower value means the output is more deterministic.

**Test Set**

A test set is a collection of sample documents representative of the challenges and types of content an ML solution will face once in production. A test set is used to measure the accuracy of an ML system after it has gone through a round of training.

**Text Analytics**

Techniques used to process large volumes of unstructured text (or text that does not have a predefined, structured format) to derive insights, patterns, and understanding; the process can include determining and classifying the subjects of texts, summarizing texts, extracting key entities from texts, and identifying the tone or sentiment of texts.

**Text Summarization**

A range of techniques that automatically produce short textual summaries representing longer or multiple texts. The principal purpose of this technology is to reduce employee time and effort required to acquire insight from content, either by signaling the value of reading the source(s), or by delivering value directly in the form of the summary.

**Thesauri**

Language or terminological resource “dictionary” describing relationships between lexical words and phrases in a formalized form of natural language(s), enabling the use of descriptions and relationships in text processing.

**Tokens**

A unit of content corresponding to a subset of a word. Tokens are processed internally by LLMs and can also be used as metrics for usage and billing.

**Tokenization**

The process of breaking text into individual words or subwords to input them into a language model. Example: Tokenizing a sentence "I am ChatGPT" into the words: “I,” “am,” “Chat,” “G,” and “PT.”

**Training data**

The collection of data used to train an AI model.

**Training Set**

A training set is the pre-tagged sample data fed to an ML algorithm for it to learn about a problem, find patterns and, ultimately, produce a model that can recognize those same patterns in future analyses.

**Transfer learning**

A technique in which a pretrained model is used as a starting point for a new ML task.

**Transformer**

A type of neural network architecture designed to process sequential data, such as text. Example: The transformer architecture is used in models like ChatGPT for natural language processing tasks.

**Treemap**

Treemaps display large amounts of hierarchically structured (tree-structured) data. The space in the visualization is split up into rectangles that are sized and ordered by a quantitative variable. The levels in the hierarchy of the treemap are visualized as rectangles containing other rectangles.

**Triple or Triplet Relations aka (Subject Action Object (SAO))**

An advanced extraction technique which identifies three items (subject, predicate and object) that can be used to store information.

**Tunable**

An AI model that can be easily configured for specific requirements. For example, by industry such as healthcare, oil and gas, departmental accounting or human resources.

**Tuning (aka Model Tuning or Fine Tuning)**

The procedure of re-training a pre-trained language model using your own custom data. The weights of the original model are updated to account for the characteristics of the domain data and the task you are interested modeling. The customization generates the most accurate outcomes and best insights.

**Turing test**

The Turing test was created by computer scientist Alan Turing to evaluate a machine’s ability to exhibit intelligence equal to humans, especially in language and behavior. When facilitating the test, a human evaluator judges conversations between a human and machine. If the evaluator cannot distinguish between responses, then the machine passes the Turing test.

**Unstructured Data**

Unstructured data do not conform to a data model and have no rigid structure. Lacking rigid constructs, unstructured data are often more representative of “real world” business information (examples – Web pages, images, videos, documents, audio).

**Unsupervised Learning**

A type of machine learning in which a model is trained on unlabeled data to find patterns or features in the data. Example: An unsupervised learning algorithm that can cluster similar images of handwritten digits based on their visual features.

**Voice Processing**

Voice processing in AI refers to the pipeline of speech-to-text conversion followed by text-to-speech synthesis.

**Windowing**

A method that uses a portion of a document as metacontext or metacontent

**Weak AI**

Weak AI refers to narrow systems that excel at specific tasks within limited contexts, but lack generalized intelligence and adaptability outside their domain.

**X-risk?**

X-risk, short for existential risk, refers to the potential for highly advanced artificial intelligence to pose an existential threat to humanity through unintended consequences or goal misalignment.

**Zero shot extraction**

The ability to extract data from text with no previous training or annotations.

**Zero-Shot Learning**

Zero-shot learning is a technique in which a machine learning model can recognize and classify new concepts without any labeled examples.

**Zero-to-One Problem**

The zero-to-one problem refers to the difficulty of finding an initial solution when addressing complex challenges, which is often disproportionately challenging compared to subsequent progress.